NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

SUBMARINE BASE, GROTON, CONN.

REPORT NUMBER 1003

EVALUATION OF A COMPUTER-ASSISTED DIAGNOSIS PROGRAM
FOR ACUTE ABDOMINAL PAIN WITH PHYSICIAN-COLLECTED DATA

by

Steven F. OSBORNE, LT, MC, USNR

Released by:

W. C. MILROY, CAPT, MC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory
26 March 1985

EVALUATION OF A COMPUTER-ASSISTED DIAGNOSIS PROGRAM
FOR ACUTE ABDOMINAL PAIN WITH PHYSICIAN-COLLECTED DATA

by

Steven F. OSBORNE, LT, MC, USNR

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

REPORT NUMBER 1003

Approved and Released by:

W. C. MILROY, CAPT, MC, USN
Commanding Officer
NavSubMedRschLab

THE PROBLEM

A computer program designed to aid the Independent Duty (8402) Corpsman in diagnosis and management of abdominal pain has been placed aboard submarines for test and evaluation. A parallel evaluation of the program's efficacy when used by physicians in the emergency room of a Naval hospital is reported here.

THE FINDINGS

Percentage agreement between initial diagnosis and final diagnosis was about the same whether the source of the initial diagnosis was physician judgment or computer calculation based on physician-supplied data. Accuracy of both sets of initial diagnoses was greatest for cases ultimately diagnosed as "non-specific abdominal pain." The relatively small number of patients otherwise afflicted precludes any firm conclusion regarding initial diagnoses in those cases.

APPLICATIONS

The program seems "medically safe" when it concludes with high probability that the patient has presented with non-specific abdominal pain.

ADMINISTRATIVE INFORMATION

Abstract

In July 1982 the Naval Submarine Medical Research Laboratory launched a five-year study aboard nuclear submarines of the computer-assisted diagnostic program for abdominal pain. Concurrently, a similar study with physician-collected data at the Emergency Room at the Naval Hospital, Groton, CT, was begun. A total of 90 cases of abdominal pain (male and female) was collected. Approximately 80% of these had a final diagnosis of non-specific abdominal pain. The program seems "medically safe" when this category is predicted with a high probability. Due to the limited number of cases collected, however, the program is not validated for this category or the other five categories of abdominal pain.

Further data collection is anticipated and desirable. Over a five-year period sufficient data should be collected to permit at least a partial validation of both systems.

EVALUATION OF A COMPUTER-ASSISTED DIAGNOSIS PROGRAM
FOR ACUTE ABDOMINAL PAIN WITH PHYSICIAN-COLLECTED DATA

I.   BACKGROUND

The Naval Submarine Medical Research Laboratory (NSMRL) is tasked with developing new methods and procedures for assisting the medical department representative (MDR) aboard deployed United States Navy submarines. Paramount among the duties of the MDR is the evaluation and treatment of crewmembers who are ill. The history and physical exam are the cornerstone of the evaluation, with laboratory facilities being modest. Consultation with other medical personnel is attempted only in emergencies and if the mission permits it. With the knowledge that computer-assisted diagnosis was under study elsewhere, NSMRL saw an opportunity to develop a computerized diagnostic aid for use by the MDR. Since abdominal pain was the most frequent chief complaint in medical evacuations from submarines, development of a computer program to assist in differential diagnosis of this condition was initiated several years ago. The intent was (and is) to provide a "sophisticated decision aid" to supplement the modest laboratory facilities and to compensate for the lack of a consultant.

A prototype program was formulated in conjunction with Dr. F. T. de Dombal of Leeds Hospital in the United Kingdom. It was modified for the submarine population and made ready for sea trials. Proper scientific and medical protocol required that the project be conducted as an experiment and that

a control (C) and experimental (E) group be formed. A protocol was approved by the Committee for the Protection of Human Subjects which involved informed consent for participation from MDR's and their randomization into a control and experimental group. The Control group has the datasheet, reference manual, and (recently) the training program. A five-year study period was projected from the anticipated case load so that statistical requirements could be met. The case load was estimated from incidence data for abdominal pain supplied by MDR's. General descriptions of the program have been previously reported.[1-5]

In January 1982, approval was granted for Fleet T&E support. In May 1982, training of Squadron Medical Officers and Squadron Corpsmen was accomplished with concomitant distribution of computer tapes and instructional materials. Individual submarine corpsmen were then trained by Squadron Medical personnel. In July 1982, sea trials were initiated. Concurrent with the sea trials, NSMRL conducted an intra-laboratory investigation of the diagnostic program using physician-collected data. The purpose was to provide an additional assessment of the efficacy of the system. It was not expected that this study would rigorously validate the database. This report summarizes the findings of that investigation.

II. METHODS

Using the abdominal pain data-sheet previously described[2], physician-students and physician-graduates of the Naval Undersea Medical Institute collected data when assigned to duty at the Naval Hospital Groton (NHGRTN) Emergency Room from May 1982-April 1984. The protocol was explained to all physicians and the datasheet definitions reviewed. Physicians were asked to complete a datasheet on all patients with a chief complaint of abdominal pain, severity unspecified. Patients old enough to relate a clear history (about age 5), and older, were included as were female patients who, in the opinion of the practitioner, had a non-gynecologic source of pain. Datasheets had to be completed preferably at the time of the visit but certainly within 48 hours. On occasion, the emergency treatment record (ETR) was used by the physician to recall details of the case. If datasheets were not completed within 48 hours, or if more than 3 of the 38 categories on the datasheet were incomplete, then the data was not used.

Approximately ten physician-graduates stationed full time in the NHGRTN area collected data. On the average, they spent 24 hours per month in the Emergency Room. Approximately 25 physician-students collected data, spending a total of 600 physician-hours in the Emergency Room. Follow-up visits and calls were periodically made to those collecting data to remind them of the study, to provide datasheets, etc. No attempt was made to double check the number of patients with abdominal pain presenting during a physician's duty hours with the number of datasheets submitted. Physicians were promised follow-up of the patient's outcome and a computer printout from the program as an inducement to collect data.

Datasheets returned to NSMRL were reviewed for completion and briefly compared with the ETR. Any discrepancies were reviewed with the physician or the patients. The ETR was considered the correct document if a discrepancy was unresolved, although there were minimal discrepancies and no major ones (i.e., on occasion the vital signs were different). Follow-up was accomplished by means of a phone call to the patient at least 7 days after the ER visit. The patient was then followed until the pain resolved or a specific diagnosis had been made. Final diagnosis was assigned by the author based on the overall record and the phone call. If the patient had been admitted to the hospital, then the discharge diagnosis was used. If an operation had been performed, the pathologist's diagnosis was used.

After assignment of final diagnosis, the data were entered into the computer and a "case summary page"[2,4] was generated. The program is written such that the probabilities of the six categories of diagnosis (Table I) total 100%. The conditional probabilities of the database were identical for males and females. The prior probabilities were different for males and females (Table II) although those for males were identical to prior probabilities used in the at-sea trials. Both sets of prior probabilities were obtained through de Dombal.[1]

Table I

Categories of Diagnosis

Diagnostic Category	Abbreviation
Appendicitis	APPY
Nonspecific Abdominal Pain	NONSAP*
Renal Colic	RCOLIC
Perforated Duodenal Ulcer	PERFDU
Cholecystitis	CHOLE
Small Bowel Obstruction	SMBOBS

*Defined as abdominal pain that is non-surgical, not life-threatening, and not requiring medical evacuation.

Table II

Prior Probabilities*

Diagnostic Category	Males	Females
Appendicitis	.18	.12
Nonspecific Abdominal Pain	.70	.75
Renal Colic	.03	.01
Perforated Duodenal Ulcer	.001**	.001**
Cholecystitis	.05	.11
Small Bowel Obstruction	.03	.02

*Rounded to nearest hundredth, except for PERFDU.

**This is an artificial number used to permit calculations while reflecting the low prior probability.

III. RESULTS

A total of 91 datasheets with supporting documentation were submitted by 10 different practitioners. Three of these practitioners accounted for 66 cases. Two cases were submitted by physicians assistants (unsolicited) and were reviewed by the supervisory medical officer in the ER. These cases were included as the data were collected properly. One case was discarded due to inadequate follow-up after a transfer. A total of 90 cases were accepted for review, 51 male cases and 39 female cases. Table III shows the case number with the practitioner's initial diagnosis, the computer diagnosis, and the final diagnosis tabulated for male cases. The computer diagnosis has a calculated probability of $\geq 95\%$ unless otherwise listed. Table IV shows similar data for female cases.

From Table III (male data), concordance between the practitioner's initial diagnosis and the final diagnosis was 88% (45/51). Concordance between the computer diagnosis and the final diagnosis was 84% (43/51). Concordance between the practitioner's initial diagnosis and the computer diagnosis was 88% (45/51). For the purpose of this comparison, the computer diagnosis was arbitrarily taken to be that category with a calculated probability greater than 50%.

In the instances where a final diagnosis of OTHER was obtained (cases 9, 23, 25, 37), both the practitioner and the computer were judged to be incorrect. While the practitioner had the option of selecting OTHER, the computer could not select this category. One might argue that for cases 9 and 23 (final diagnoses infectious mononucleosis and inguinal hernia), NONSAP was appropriate for practitioner and computer. In case 25 (final diagnosis - pneumonia), the presentation was unusual. The management of pneumonia is clearly different from that of NONSAP. In case 37 (final diagnosis - mesenteric adenitis), the distinction between APPY and NONSAP is difficult. In a sense, the computer's calculation of APPY - 58%, NONSAP - 41% is keener than the practitioner's selection of APPY.

Of note is the high percentage of cases with a final diagnosis of NONSAP (41/53, 77% of total cases). This is to be expected as NONSAP has been found to have a prior probability of 80% in ER settings. If these cases are taken alone, concordance between the practitioner's diagnosis and the final diagnosis is 97% (40/41). Similarly, concordance between the computer diagnosis and the final diagnosis is 90% (37/41). Concordance between the practitioner's diagnosis and the computer diagnosis was 88% (36/41).

From Table IV (female data) concordance between the practitioner's initial diagnosis and the final diagnosis was 85% (33/39). Concordance between the computer diagnosis and the final diagnosis was 85% (33/39). Concordance between the practitioner's initial diagnosis and the computer diagnosis was 69% (27/39). The computer program's selection with a probability greater than 50% was again taken as the computer diagnosis. Case 16 left the computer "in error" as it calculated NONSAP 50%, APPY 49%. Cases 2 and 18 involved pregnant women, although it was felt by the practitioners that the cause of pain

Table III

USNHGROTON CASE DATA: ABDOMINAL PAIN – MALE

CASE #	INITIAL D_X	COMPUTER D_X	FINAL D_X
1	NONSAP	APPEND-57%;NONSAP-41%	NONSAP
2	NONSAP	NONSAP	NONSAP
3	NONSAP	NONSAP	NONSAP
4	SMBOBS	RCOLIC-86%;NONSAP-13%	RENAL COLIC
5	NONSAP	NONSAP	NONSAP
6	APPY	NONSAP-86%;APPY-13%	NONSAP
7	NONSAP	NONSAP	NONSAP
8	NONSAP	NONSAP	NONSAP
9	NONSAP	NONSAP	OTHER (Infectious Mononucleosis)
10	NONSAP	APPY-94%;NONSAP-5%	NONSAP
11	NONSAP	SMBOBS	NONSAP
12	NONSAP	NONSAP	NONSAP
13	NONSAP	NONSAP	NONSAP
14	NONSAP	NONSAP	NONSAP
15	NONSAP	NONSAP	NONSAP
16	NONSAP	NONSAP	NONSAP
17	NONSAP	NONSAP	NONSAP
18	NONSAP	NONSAP	NONSAP
19	RENAL COLIC	RENAL COLIC	RENAL COLIC
20	NONSAP	NONSAP-87%;APPY-12%	NONSAP
21	NONSAP	NONSAP-83%;APPY-16%	NONSAP
22	NONSAP	NONSAP	NONSAP
23	NONSAP	NONSAP-75%;APPY-24%	OTHER (Inguinal Hernia)
24	APPY	APPY	APPY
25	NONSAP	NONSAP	OTHER (Pneumonia)
26	NONSAP	NONSAP	NONSAP
27	NONSAP	NONSAP	NONSAP
28	NONSAP	NONSAP	NONSAP
29	NONSAP	NONSAP-85%;APPY-12%	NONSAP
30	NONSAP	NONSAP	NONSAP
31	NONSAP	NONSAP-75%;APPY-24%	NONSAP
32	NONSAP	NONSAP	NONSAP
33	NONSAP	NONSAP	NONSAP
34	NONSAP	NONSAP	NONSAP
35	NONSAP	NONSAP	NONSAP
36	NONSAP	NONSAP	NONSAP
37	APPY	APPY-85%;NONSAP-41%	OTHER (Mesenteric Adenitis)
38	NONSAP	NONSAP	NONSAP
39	RENAL COLIC	RCOLIC-52%;NONSAP-31%	RENAL COLIC
40	NONSAP	NONSAP	NONSAP

Table III Cont'd

41	APPY	APPY-89%;NONSAP-10%	APPY
42	NONSAP	NONSAP	NONSAP
43	NONSAP	NONSAP	NONSAP
44	NONSAP	NONSAP	NONSAP
45	NONSAP	NONSAP	NONSAP
46	NONSAP	NONSAP	NONSAP
47	NONSAP	NONSAP-90%;RCOLIC-9%	NONSAP
48	APPY	APPY-91%;NONSAP-8%	APPY
49	NONSAP	NONSAP	NONSAP
50	NONSAP	APPY-66%;NONSAP-33%	NONSAP
51	NONSAP	NONSAP	NONSAP

*Computer D_X is 95% or greater unless otherwise noted.

Table IV

USNHGROTON CASE DATA: ABDOMINAL PAIN - FEMALE

CASE #	INITIAL D_X	COMPUTER D_X*	FINAL D_X
1	NONSAP	NONSAP	NONSAP
2	CHOLECYSTITIS/ PREGNANCY	NONSAP	"NONSAP"-PREGNANCY
3	NONSAP	NONSAP-67%;CHOLE-30%	NONSAP
4	NONSAP	NONSAP	NONSAP
5	CHOLECYSTITIS	NONSAP-51%;CHOLE-48%	CHOLECYSTITIS
6	PYELONEPHRITIS	CHOLECYSTITIS	PYELONEPHRITIS
7	NONSAP	NONSAP	NONSAP
8	NONSAP	NONSAP	NONSAP
9	APPY/ADENITIS	NONSAP-80%;APPY-19%	NONSAP
10	UTI	NONSAP	NONSAP
11	NONSAP	NONSAP	NONSAP
12	NONSAP	NONSAP	NONSAP
13	NONSAP	NONSAP	NONSAP
14	NONSAP	NONSAP	NONSAP
15	CHOLECYSTITIS	CHOLECYSTITIS	CHOLECYSTITIS
16	NONSAP	NONSAP-50%;APPY-49%	NONSAP
17	RCOLIC OR PART.SMBOBS	NONSAP	UTI
18	CHOLE	NONSAP	"NONSAP"-PREGNANCY
19	NONSAP	APPY	APPY
20	NONSAP	CHOLE-90%;NONSAP-9%	NONSAP
21	BILIARY COLIC VS. CHOLECYSTITIS	CHOLECYSTITIS	CHOLECYSTITIS
22	NONSAP	NONSAP-74%;APPY-25%	NONSAP
23	NONSAP	CHOLE-75%;NONSAP-24%	NONSAP
24	BILIARY COLIC	NONSAP	NONSAP
25	NONSAP	NONSAP	NONSAP
26	UTI	NONSAP	NONSAP
27	NONSAP	NONSAP	NONSAP
28	LG. BOWEL OBS.	NONSAP	NONSAP
29	BILIARY COLIC	NONSAP-75%;CHOLE-24%	BILIARY COLIC
30	DIVERTICULITIS	NONSAP	NONSAP
31	NONSAP	NONSAP	NONSAP
32	NONSAP	NONSAP	NON-CYSTIC OVARIAN MASS
33	NONSAP	NONSAP	NONSAP
34	NONSAP	NONSAP	NONSAP
35	NONSAP	NONSAP	NONSAP
36	NONSAP	NONSAP-86%;CHOLE-13%	NONSAP
37	NONSAP	NONSAP	NONSAP
38	NONSAP	NONSAP	NONSAP
39	NONSAP	NONSAP	NONSAP

*Computer D_X is 95% or greater unless otherwise noted.

was unrelated to pregnancy.
Possibly, these two cases should be
excluded in the final analysis.

As with the male case data,
the final diagnosis of NONSAP
accounted for a majority of the
cases - 79% (31/39). Comparisons
are similar to those above if this
category is reviewed separately.

IV. DISCUSSION

There are an insufficient
number of cases to validate the
accuracy of the diagnostics as a
whole or any of the individual
diagnostic categories. An estimate
has been made that 70 cases per
diagnostic category, or a total of
420 cases overall, would be required
for validation.[6]

Male (51 cases) and female (39
cases) databases should be
validated separately. While both
databases share the same conditional
probabilities, the prior
probabilities differ (see Table II).
These facts make a separate
validation for each seem advisable.
Given that the conditional
probabilities were derived from all
male cases, it may seem inappropriate
to be evaluating female cases. This
point is not disputed, but it is an
additional reason to perform
separate male/female validations.
If an acceptable diagnostic
accuracy were found using "male"
conditional probabilities and
female prior probabilities, this
author would accept the results.
The key is to accept for study
female patients who, in the initial
estimate of the practitioner, have
a non-gynecologic source of pain.

The accuracy of the program
seems good when the final diagnosis
is NONSAP and when the computer
predicts NONSAP. The program is both
specific and sensitive in this regard.
The number of cases when either
the final diagnosis or computer
diagnosis is NONSAP is sufficient
(40+) to conclude that the program
is medically "safe" for this
category. It is not stastically
validated though, and thus has not
passed rigorous scrutiny yet.

The program may "perform"
slightly better with male case data
than with female case data. This is
possibly seen in the comparison of
the practitioner's initial diagnosis
and the computer diagnosis. With
male cases this was 88% (45/51)
while with female cases it was 69%
(27/39). Although not reflected
in comparisons with the final
diagnosis, the program has a
disquieting effect when there is
more frequent disagreement with the
practitioner's diagnosis.

V. ACKNOWLEDGEMENTS

VI. REFERENCES

1. de Dombal, F. T.: Modification
    of the de Dombal computer-based
    system for diagnosis of abdominal
    pain to optimize its accuracy
    when applied to the United
    States Naval Submarine Force
    population. U.S. Navy Contract
    #N68171-79-C-9602, 1980.

2. Arthur, D. C.: Computer-assisted
    diagnosis of acute abdominal pain
    for Submarine Corspmen. U.S.

Navy Medicine, 1982, 73, 14-21.

3. Henderson, J. V., Moeller, G.,
   Ryack, B. L., & Shumaik, G.:
   Adaptation of a computer-
   assisted program for use by
   hospital corpsmen aboard
   nuclear submarines. Second
   Annual Symposium on Computer
   Applications in Medical Care,
   Washington, DC, 1978.

4. Osborne, S. F.: Medical diagnosis
   aboard submarines: Use of a
   computer-based Bayesian method
   of analysis in an abdominal pain
   diagnostic program. JOM, in press.

5. Osborne, S. F.: Computer-assisted
   diagnosis program for acute
   abdominal pain - Interim Report
   July 1982-Sept 1983. NSMRL
   Report No. 1012, Dec 1983.

6. Letter and personal communication
   between Dr. Adrian Ostfeld,
   Yale University School of
   Medicine, Dec 1982.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NSMRL Report No. 1003	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* EVALUATION OF A COMPUTER-ASSISTED DIAGNOSIS PROGRAM FOR ACUTE ABDOMINAL PAIN WITH PHYSICIAN-COLLECTED DATA		5. TYPE OF REPORT & PERIOD COVERED Interim report
		6. PERFORMING ORG. REPORT NUMBER Rpt 1003
7. AUTHOR(s) STEVEN F. OSBORNE, LT, MSC, USNR		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Submarine Medical Research Laboratory Box 900 Naval Submarine Base New London Groton, CT 06349		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS M0095.001-1045
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command Naval Medical Command, National Capital Region Bethesda, MD 20814		12. REPORT DATE 26 March 1985
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Abdominal pain; computer-assisted diagnosis

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

In July 1982 the Naval Submarine Medical Research Laboratory launched a five year study aboard nuclear submarines of the computer-assisted diagnostic program for abdominal pain. Concurrently, a similar study with physician-collected data at the Emergency Room at the Naval Hospital, Groton, CT, was begun. A total of 90 cases of abdominal pain (male and female) was collected. Approximately 80% of these had a final diagnosis of non-specific abdominal pain. The program seems "medically safe" when this category is predicted with

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-014-6601 |

Item 20 -- continued

a high probability. Due to the limited number of cases collected, however, the program is not validated for this category or the other five categories of abdominal pain.

Further data collection is anticipated and desirable. Over a five year period, sufficient data should be collected to permit at least a partial validation of both systems.